DOCUMENT RESUME

ED 194 829 CG 014 770

AUTHOR Gill, Roger W.
TITLE What is the Content in "Content Validity"?
PUB DATE 80
NOTE 12p.: Paper presented at the Annual Convention of the American Psychological Association (88th, Montreal, Quebec, Canada, September 1-5, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavioral Science Research: *Evaluation Criteria: Foreign Countries: Job Analysis: *Job Performance: Predictive Validity: Research Methodology: Research Problems: *Skill Analysis: *Task Analysis: *Work Sample Tests
IDENTIFIERS *Content Validity: *England

ABSTRACT

Although predictive validation is the primary method of justifying a particular work sample for use as an assessment device, content validation is being recognized as an alternative strategy. Confusion exists over whether content validity should refer to the representativeness of the work sample, in terms of tasks to be performed, or behavioral aspects of work performance. The in-basket exercise, a simulation of the typical contents of a manager's in-basket which require decision and action, is the work sample used with 18 lower- and middle-level managers. Effects were investigated of two in-basket exercises based upon task sampling one of which was job-related to the subjects, and two of which used job-related criteria. Results suggest that job-relatedness of assessment criteria is more important than job-relatedness of the in-basket exercise, implying a need to redefine content validity. Content validity is implied to relate to behavioral aspects of work performance rather than to the task content. (Author/NRB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194829

Paper to APA Division 14

1980 Annual Convention

WHAT IS THE CONTENT IN "CONTENT VALIDITY"?

Roger W. T. Gill

School of Management
State University of New York at Binghamton
Binghamton, New York 13901

CG014770

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roger W. T. Gill

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

WHAT IS THE CONTENT IN "CONTENT VALIDITY"?

by
Roger W. T. Gill
School of Management
State University of New York at Binghamton

SUMMARY

Statement of Problem

The assessment of performance in work sample exercises is beset by the problem of validation. Work samples have usually been thought of as best realistically sampling or simulating the work performed in the job (Wernimont and Campbell, 1968; Standards and Ethical Considerations for Assessment Center Operations, 1979). Whilst predictive validation has remained the primary method, content validation has been increasingly recognized as an alternative, or at least interim, acceptable strategy for justifying a particular work sample for use as an assessment device (Gorham, 1978), although there is considerable controversy over whether this is sound practice (Tenopyr, 1977; Guion, 1977, 1978).

Furthermore, the nature of content validity is by no means clear: the Courts have failed to act upon a uniform set of standards for determining content validity (Kleiman and Faley, 1978). In particular, there is confusion over whether content validity should refer to the representativeness of the work sample in terms of the tasks to be performed or problems encountered in the real job or to the representativeness of the "skills, knowledges or behaviors.....as..... samples of the content of the target jobs" (Taylor, 1978). In the former case, content refers to task; in the latter case, to behavior. In short, what precisely should be the content in "content validity"?

What underlies both approaches, of course, is the need for thorough job analysis. However, where a work sample has maximal content validity in terms of its sampling of the tasks or work to be done, if the criteria used to assess subjects' performance are unrepresentative of how job-holders are perceived to be effective performers, then

content validity so defined is virtually useless in practice. On the other hand, assessing the performance of subjects in a work sample exercise which does not closely resemble the job in question, but using assessment criteria which are known to be significantly related to effective performance in the job, may yield more accurate and more fair measures of likely performance in the job.

Current methods of establishing content validity therefore may be criticized. For example, Lawshe's (1975) procedure does not identify or describe the actual skill or knowledge which a judge construes as measured by an item in the work sample exercise and which is judged to be essential (or useful or not necessary) to the performance of the job in question. There needs to be a reasonable degree of agreement that behaviors of a certain kind, namely empirically identified and explicitly described criterion behaviors, are elicited by the work sample exercise.

This paper reports an investigation of the hypothesis that a more content-valid work sample exercise permits significantly greater demonstration of criterion behaviors than a less content-valid exercise, content validity being taken to refer to situational specificity and task representativeness and established by Lawshe's procedure.

Procedure

Subjects

The subjects were 18 lower-and middle-level managers and supervisors in a glass-container manufacturing company in England.

Experimental Tool

The work sample employed was the in-basket exercise, which is a simulation of the typical contents of a manager's in-basket, comprising correspondence, reports, messages, and so on, which require decisions and action. This exercise has been widely used in the assessment and development of managers, with a high degree of acceptability and with relatively good empirical validities (Gill, 1979).

Method

Two in-basket exercises were developed, one specifically for a managerial job in

the glass-container manufacturing company and the other specifically for a local government officer's job in a typical British metropolitan district council, using a sampling process in respect of a number of in-baskets of actual job-incumbents.

Behavioral assessment criteria were determined for both jobs in order to permit a comparison of the extent to which the criterion behaviors were demonstrated by the group of 18 subjects in the two, differentially content-valid in-basket exercises.

This was accomplished for the industrial in-basket by the "brainstorming" and group discussion approach described by Gough (1970) in respect of a group of 18 main board directors, executives and senior managers, leading to a list of 10 behavioral criteria of effectiveness of the industrial manager. For the local government situation, however, there was time and opportunity to use a more sophisticated method to reduce bias in perception and reporting, the interview procedure in repertory grid technique, based upon the theory of personal constructs (Stewart and Stewart, 1976; Bannister, 1969; Easterby-Smith, 1976). This involved an interview with 50 (65%) officers in the target job to identify the ways in which they construe their jobs, using the "minimum context card form" (Kelly, 1955) in which they were asked firstly to identify eight work activities (elements) which are important, occupy much time or are done frequently, and, though important, are unexpected or emergency kinds of activity, and then to say what characterized any pair of elements in each of a series of randomly presented triads. Behavioral exmaples of those constructs were next generated and a questionnaire devised consisting of a series of bipolar or "opposite" statements based upon the behavioral examples.

The performance questionnaire was distributed to the 13 bosses of the 50 officers who were asked to rate their most effective (anonymous) subordinate on the behavioral criteria. Some weeks later, this process was repeated in respect of their least effective subordinate.

The frequency of responses was scored for all items and extreme scores statistically weighted. By analyzing discriminability of the items, separation scores were

computed which showed those 32 items (behavioral assessment criteria) which significantly distinguished between effective and ineffective officers. Twenty of these behaviors were judged to be measurable in an in-basket exercise as the behavioral criteria of effectiveness of the local government officers.

In order to check the respective content validities of the two in-basket exercises, Lawshe's (1975) procedure was adopted. Twelve MBA students at an American university were familiarized in depth with the two organizations. Next they performed both in-basket exercises. Each student (assessor) then was asked whether the skill or knowledge measured by each item in the in-basket was (a) essential, (b) useful, or (c) not necessary, to the performance of the job. This was done for both in-baskets in respect of both jobs. Table 1 shows the results of content validation using this procedure.

A further four assessors joined the 12 assessors described above and, following individual performance of both in-basket exercises, all were trained in behavior rating, specifically in the categorizing of observed behavior of subjects using the 30 assessment criteria (10 from industry and 20 from local government). Three completed in-basket exercises, tackled by different subjects, were rated by each of the assessors, with duplication in eight cases to check interrater reliability.

The extent to which criterion behaviors were demonstrated was analyzed by a two-way analysis of variance based upon a repeated-measures, two-factors design.

Results

The results of the content evaluation by the assessors are shown in Table 1 below.

Table 1 about here

The t tests indicate that both in-basket exercises were significantly different from each other in terms of average item relevance respectively. In other words, each was content valid in respect of the job situation to which it referred.

The inter-rater correlation coefficient (for eight pairs of assessors, i.e. 6,480 pairs of ratings) was 0.46 ($p \leq 0.001$).

Table 2 shows the overall average ratings for the 18 Ss on both in-baskets using both sets of criteria.

Table 2 about here

Analysis of variance on these data yielded the results shown in Table 3.

Table 3 about here

It is clear that the nature of the in-basket exercise, whether it was "content-valid" or not, had no significant effect on the extent to which criterion behaviors were demonstrated. However, the particular set of criteria used had a very significant effect on the extent to which the behaviors thus measured were demonstrated. The effects of the nature of the criteria used and the nature of the in-basket interacted to a significant degree.

Discussion

The experimental hypothesis was not supported. The implications of the findings underscore the importance of defining criteria on which assessments are to be made rather than the need to develop a content-valid in-basket exercise in terms of task sampling or situational specificity. The use of the industrial criteria permitted more industrial criterion behaviors to be demonstrated with the use of both the industrial and the local government in-basket exercise than did the use of the local government criteria. Since the subject group was familiar with the industrial in-basket exercise, this finding might have been expected on the basis. However, since the industrial criteria applied to the local government in-basket exercise also allowed more of the industrial criterion behaviors to be demonstrated than did the local government in-basket exercise, despite the respective content validity of both in-basket exercises,

it appears that what is particularly important is the relevance and the specificity of the assessment criteria rather than the content validity of the in-basket exercise itself in terms of the situation portrayed or the tasks sampled.

This helps us to answer the question of the nature of content validity. It was suggested earlier that confusion exists over whether content validity relates to the task content or to the behavioral content of a work sample exercise. From this study, content validity is implied to relate to the skill or behavioral aspects of work performance rather than to the task or work content. In accord with recent legal decisions is the need implied by these findings for adequate job analysis leading not only to proper job descriptions but also to accurate statements of criteria of effectiveness or success by which to judge candidates in selection, training or promotion circumstances. This study emphasizes the importance of having specific and accurate criteria of effectiveness by which to judge subjects.

Methodological and analytical shortcomings may be identified as consisting in the use of two, different methodologies for determining criteria of effectiveness, the averaging of measures of criterion behaviors across all 20 local government criteria and all 10 industrial criteria, and the marginal, though significantly high, interrater reliability. Nevertheless, the findings support the latest APA Division 14 Draft Revision of the Principles for the Validation and Use of Personnel Selection Procedures (1979) in pointing to the relevant content domain as performance (rather than task).

TABLE 1

Content Evaluation[1] of Industrial and Local Government In-Basket Exercises

Local Government In-Basket Average Item Relevance to:		Industrial In-Basket Average Item Relevance to:	
Local Government Officer (A)	Industrial Manager (B)	Local Government Officer (C)	Industrial Manager (D)
8.39^{2} σ=1.63	5.23^{2} σ=2.16	5.68^{2} σ=1.81	6.75^{2} σ=2.96

Notes:

[1] Figures stated are the mean number of judges rating the skill or knowledge measured by the items as "essential"

[2] $t_{A,B} = 8.41$ $(p \leq 0.001)$

$t_{C,D} = 2.39$ $(p \leq 0.02)$

$t_{A,C} = 5.63$ $(p \leq 0.001)$

$t_{B,D} = 2.98$ $(p \leq 0.005)$

TABLE 2

Subjects' Overall Average Ratings[1] on Two In-Basket Exercises Using Two Sets of Assessment Criteria

Subject	Industrial In-Basket		Local Government In-Basket	
	Local Govt. Criteria	Industrial Criteria	Local Govt. Criteria	Industrial Criteria
1	0.31	0.53	0.31	0.63
2	0.40	1.04	0.51	0.93
3	0.47	0.60	0.53	0.94
4	0.72	1.51	0.62	1.15
5	1.05	1.60	0.46	0.71
6	0.19	0.48	0.86	1.09
7	0.48	1.04	0.28	0.31
8	0.74	1.31	1.27	2.18
9	1.61	2.59	0.33	0.99
10	0.54	1.04	0.41	0.77
11	0.41	0.87	0.34	0.47
12	1.44	1.88	0.25	0.52
13	0.76	1.11	1.37	2.17
14	0.92	1.15	1.24	1.66
15	0.73	1.18	1.34	1.44
16	0.23	0.62	0.68	1.32
17	1.45	2.57	0.52	0.97
18	1.36	2.15	0.77	2.01
$\bar{x}$	0.77	1.29	0.67	1.12
s	0.45	0.65	0.39	0.57

Note:

[1]Ratings or the basis of 1 (low) to 5 (high) have been rounded so total averages may appear to be understated.

TABLE 3

Analysis of Variance of Subjects' Overall Average Ratings on Two In-Basket Exercises using Two Sets of Assessment Criteria

Source of Variation	SS	df	MS	F	p
Total	22.75	71	-	-	-
Subjects	8.89	17	-	-	-
In-Basket Exercises	0.31	1	0.31	0.63	N.S.
Assessment Criteria	4.32	1	4.32	81.54	$\leq$0.001
Assessment Criteria x In-Basket Exercises	0.02	1	0.02	5.64	$\leq$0.05
Error (In-Basket Exercises)	8.38	17	0.49	-	-
Error (Assessment Criteria)	0.90	17	0.05	-	-
Error (Assessment Criteria x In-Basket Exercises)	0.07	17	0.004	-	-

Key: SS = Sum of squares

df = degrees of freedom

MS = mean square

f = Ratio of MS of the main source of variation to MS of error variation

p = probability level

References

APA Division 14 (1979). Principles for the Validation and Use of Personnel Selection Procedures. Draft Revision.

Bannister, D. (1969). The Application of Repertory Grid Technique in Occupational Psychology. Third Occupational Psychology Conference of the British Psychological Society, January.

Easterby-Smith, M. (1976). The Repertory Grid as a Personnel Tool. Management Decision. 14, (5), 239-247.

Gill, R. W. T. (1979). The In-Tray (In-Basket) Exercise as a Measure of Management Potential. Journal of Occupational Psychology, 52, (3), 185-197.

Gorham. W. A. (1978). Federal Executive Agency Guidelines and their Impact on the Assessment Center Method. Journal of Assessment Center Technology. 1, (1), 2-8

Gough, J. S. (1970). Recruitment of Graduates at the Universities, in B. Ungerson (ed.), Recruitment Handbook, London: Gower Press.

Guion, R. M. (1977). Content Validity: Three Years of Talk - What's the Action? Public Personnel Management. November/December, 407-414.

Guion, R. M. (1978). A Note on Taylor's "EEOC Guidelines on Content Validity". Journal of Assessment Center Technology. 1, (3), 15-17.

Kelly, G. A. (1955). The Psychology of Personal Constructs. Norton.

Kleiman, L. S. and Faley, R. H. (1978). Assessing Content Validity: Standards Set by the Courts. Personnel Psychology. 31, 701-713.

Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. Personnel Psychology. 28. 563-575.

Standards and Ethical Considerations for Assessment Center Operations (1979), J. L. Moses (Chairman).

Stewart, A. and Stewart, V. (1976). Tomorrow's Men Today, London: Institute of Personnel Management.

Taylor, J. O. (1978). The E.E.O.C. Guidelines on Content Validity and the application to the Assessment Center Method. Journal of Assessment Center Technology. 1, (1), 8-12.

Tenopyr, M. L. (1977). Content-Construct Confusion. Personnel Psychology. 30. 47-54.

Wernimont, P. F. and Campbell, J. P. (1968). Signs, Samples, and Criteria. Journal of Applied Psychology. 52. 372-376.